A NEW SEMICONDUCTOR TETRODE, THE SURFACE-POTENTIAL CONTROLLED TRANSISTOR

C. T. Sah

Fairchild Semiconductor
A Division of Fairchild Camera and
Instrument Corporation
844 Charleston Road
Palo Alto, California

32/2

1961 / WESTERN ELECTRONIC SHOW AND CONVENTION
San Francisco, August 22-23-24-25, 1961

A NEW SEMICONDUCTOR TETRODE, THE SURFACE-POTENTIAL CONTROLLED TRANSISTOR

C. T. Sah
Fairchild Semiconductor
A Division of Fairchild Camera & Instrument Corporation
844 Charleston Road
Palo Alto, California

Summary

The theory and operating characteristics of a new semiconductor tetrode is discussed in this paper. This semiconductor junction device has the usual geometry of the planar transistor but with an additional metal electrode placed on the oxide which covers the surface of the emitter-base junction. This electrode serves as a grid. The grid-base voltage controls the surface potential, surface recombination rate and the size of the surface channel, and thereby the current gain of the transistor. In the common emitter connection, the grid serves as a second input in addition to the base input. Input impedance of the grid is typically in the range of 1 to 100 pf and 10^{15} ohms. The transconductance, dI_{ce}/dV_{ge}, of several thousand micromhos has been achieved which is essentially limited in frequency response by the h_{fe} cutoff due to transit time in the transistor.

Introduction

It is well known that surface plays a very important part in controlling the electrical characteristics of transistors and diodes. It has been demonstrated recently that surface protection by thermal oxidation of silicon transistor (the planar transistor) greatly improves the low current characteristics such as the leakage current and the current gain factor, h_{FE}.[1] In addition, oxide surface protection has provided orders of magnitude improvement in the stability and reliability in the overall performance of transistors and diodes.

The advancement of the planar technology in the semiconductor art has not only improved the understanding of the electrical characteristics of transistors and diodes through controlled and designed experiments, but has also made it possible to design and fabricate more sophisticated active and passive semiconductor components. For example, the planar technique has made it possible to fabricate the micrologic element or integrated circuit which consists of transistors, diodes, resistors and capacitors all built on the same block of silicon by solid state diffusion techniques. In this paper, one of a new family of semiconductor devices will be described which features an almost open circuit input impedance.

This family of semiconductor junction devices consists of modified versions of the usual diode, triode and pnpn or multijunction switches. In addition to the usual electrodes, these devices have one or more metal electrodes placed on top of the oxides which cover the surface of the junction transition regions. These metal electrodes, or the grids, control the surface potential, the recombination rate, and the size of the surface channel. For example, in the one junction version, the grid voltage may modulate the current-voltage characteristics of the junction through the creation or elimination and control of the size of the channels which are connected to the transition region of the junction. The reverse (i.e. p-side negative-biased with respect to the n-side) characteristics are quite similar to the field effect transistor. However, there is a major difference: in the surface-potential controlled devices, minority carrier recombination and generation in the channel provides a considerable amount of junction current while in the field effect transistors, the modulation by the field electrode is obtained through the control of the channel conductance of the majority carrier.[2] In this paper, we shall discuss the electrical performance and physical theory of the tetrode which has two junctions or three layers with one field electrode.

Electrical Characteristics

A cross-sectional view of the tetrode is shown in Fig. 1. Both the emitter-base and the collector-base junctions are oxide protected. A grid electrode is placed on the oxide of the emitter-base junction surface.

h_{FE} - I_{CE}

The d.c. beta or h_{FE} is measured as a function of collector current in the common emitter configuration with the grid voltage as a parameter. The experimental data are plotted for the high voltage version of the npn and pnp units in Fig. 2 and 3. The leakage current, I_{CEO}, is less than 0.5mua for both units. The characteristics are very similar and show considerable reduction of current gain at low current and large grid voltage. The slope of the

curves asymtotically approach 0.5 at large grid voltage and low currents, i.e. $h_{FE} = (I_C/I_B) \sim \sqrt{I_C}$, indicating that the base current varies approximately as $\sqrt{I_C}$. It has been observed that the collector current varies as exp. (qV_{EB}/kT) over almost ten decades of collector current, thus the base current varies as exp $(qV_{EB}/2kT)$.

I_{CE} - V_{GE}, (Transconductance)

The transfer characteristics are also obtained in the grounded emitter configuration and are shown in Fig. 4, 5 and 6. The tetrode, 1NP2 of Fig. 4 is a low grid voltage NPN device which has thinner oxide and lower surface concentration in the base region than the NPN tetrode, 1W1-3 shown in Fig. 5.

For the low voltage tetrode, 1NP2, shown in Fig. 4, the maximum h_{FE} for a given base current is obtained at a negative grid voltage of approximately -8 volts at low currents and the optimum grid voltage decreases to -12 volts at high current. The peak h_{FE} grid voltage varies in a similar way for the high grid voltage device 1W1-3 shown in Fig. 5.

Maximum transconductance, $g_m = dI_{CE}/dV_{GE}$ at a given base current, is obtained at positive grid voltage as indicated in Figs. 4 and 5. The transconductance also increases with base or collector current, and is nearly proportional to the base current or the square root of the collector current at low currents. At large base current, unit 1NP2 in Fig. 4 shows a g_m of 7000 micromhos at I_{CE} = 50 ma and gate voltage of 5v. The sign of the transconductance is negative in the positive grid voltage region. However, at large negative grid voltage, a positive transconductance is obtained. The magnitude of the negative grid voltage for large positive g_m approaches the destructive dielectric breakdown condition in the oxide.

An entirely similar behavior with the same sign for the grid voltage is observed for the transfer characteristics of pnp tetrodes, shown in Fig. 6. Thus, a positive g_m is obtained for the pnp devices with a positive grid voltage which is safely below the dielectric breakdown limit. It is believed that surface channel formation may account for the observed behavior of both the npn and the pnp devices. In the npn tetrode, the maximum g_m region of positive grid voltage corresponds to an increase in size of the channel formed on the surface of the p-type base under the grid metal electrode, while for the pnp tetrode, the channel is also formed in the p-type region but in the emitter. A more detailed discussion of the physical theory will be given in a subsequent section.

g_m-frequency

The frequency response of the transconductance appears to be limited by the usual minority carrier transit time in the base. A preliminary measurement for an npn tetrode is shown in Fig. 7. The 3 db point for g_m is about 600kc, which is nearly equal to the h_{fe} cutoff frequency of this device.

Grid Input Impedance

The input capacitance of the grid to the emitter and base lead is about 4.5pf for the high grid voltage devices, 1W1-1, 1W1-3, 47-SCT1-1 and 47-2C1-5; and about 22.5pf for the low grid voltage unit, 1NP2. The capacitance is independent of the grid voltage over the entire range up to the dielectric breakdown voltage corresponding to about 2 to 5×10^6 volts/cm.

The leakage or the input resistance has not been measured successfully with commerical electrometers indicating a value greater than 10^{14} ohms. An alternative approach is made by measuring the RC discharge time of the input capacitance using the collector current as an indicator of the grid voltage under constant base current drive and common emitter connection. The decay time constant thus measured is from one to three hours giving a leakage resistance of greater than 10^{15} ohms for the 1W1-3 unit. The variation of the experimental results of the decay time constant appear to be associated with the humidity of the room ambient and the leakage in the lead through the glass of the transistor package.

Base Input Noise

The input noise voltage is measured as a function of frequency with the grid voltage as a parameter. In Fig. 8, the noise voltage referred to the input is plotted as a function of frequency. At large positive grid voltage, the current gain is decreased but is sufficient so that the noise from the amplifier is still unimportant. The data in Fig. 8 shows that the 1/f noise is increased as much as 30db at low frequencies by the grid voltage and the cross-over frequency to shot noise, f_{co}, is increased from 700cps to 30kc when the grid voltage is increased from 0 to 10 volts. The tetrode unit shown in Fig. 8 is similar to the unit 1NP2 shown in Fig. 4. The result of the base input noise is not surprising since it is well known that in semiconductor devices the 1/f noise is closely associated with the formation of surface channels.[3] The grid input noise at present is not directly measured although it can be estimated from the base input noise.

Physical Theory

The design theory and the basic mechanisms which control the electrical characteristics of the surface-potential controlled transistors and diodes may be best analyzed using the physical circuit of the device shown in Fig. 9 (a). An enlarged view of the region where the emitter-base junction intercepts the surface is shown and the junction transition regions are shaded in Fig. 9(a). These transition regions are defined in the sense that negligible voltage drop occurs out side these regions, i.e. about kT/q or less.[4,5] The shape of the shaded transition region near the surface corresponds to a positive voltage applied to the grid electrode relative to the base electrode, thereby inducing an n-type surface channel on the base side. This situation corresponds closely to that in the npn tetrodes shown in Figs. 2, 3, 4 and 6 at fairly large positive grid voltage. The potential energy diagram showing an equi-potential surface for electrons is given in Fig. 10.

The equivalent circuit of the tetrode may be derived from Fig. 9(a) and is given in Fig. 9(b). The effect of the grid induced emitter-base surface channel may be conveniently included as a surface diode, SD, connected to the external base and emitter leads of a real silicon transistor which has no surface leakage in the emitter-base junction. Thus, the characteristics for a typical tetrode may be discussed in two parts with little interaction between the two except through the two external lead connections.

The Transistor Characteristics

In addition to the usual diffusion currents, the ideal transistor shown in Fig. 9(b) includes also the recombination and generation currents in the transition region of the junctions.[6] Thus, the usual transistor equations in the grounded base connection are (PNP):

$$J_E = j_{ET} 2\sinh\theta_E/2 + j_D[(e^{\theta_E}-1)/\gamma_N - (e^{\theta_C}-1)\alpha_T] \quad (1)$$

$$J_C = j_{CT} 2\sinh\theta_C/2 + j_D[(e^{\theta_C}-1)/\gamma_I - (e^{\theta_E}-1)\alpha_T] \quad (2)$$

where

$$\theta_E = qV_{EB}/kT$$
$$\theta_C = qV_{CB}/kT$$
$$\alpha_T = \mathrm{sech}(W_B/L_B)$$
$$\gamma_N = j_D/(j_E + j_D)$$
$$\gamma_I = j_D/(j_C + j_D)$$
$$j_D = (qn_B L_B/\tau_B)\,\mathrm{ctnh}(W_B/L_B)$$
$$j_C = (qn_C L_C/\tau_C)\,\mathrm{ctnh}(W_C/L_C)$$
$$j_E = (qn_E L_E/\tau_E)\,\mathrm{ctnh}(W_E/L_E)$$
$$j_{ET} = qn_i W_{ET} f(b_E)/\sqrt{\tau_{po}\tau_{no}}\,\alpha_E$$
$$j_{CT} = qn_i W_{CT} f(b_C)/\sqrt{\tau_{po}\tau_{no}}\,\alpha_C$$

for unit emitter and collector area.

The widths W_B, W_C, and W_E are the widths of the base, collector and emitter layers respectively, while W_{CT} and W_{ET} are the widths of the collector and emitter transition regions. The n's are the minority carrier concentrations and the τ's are the minority carrier lifetime in the emitter, collector and base layers, while n_i is the intrinsic carrier concentration. The L's are the minority carrier diffusion lengths given by $L = \sqrt{D\tau}$, and τ_{po} and τ_{no} are the minority carrier lifetimes in heavily doped n-type and p-type specimen respectively. The quantity $f(b)/\alpha$ is given in reference 4 and may be approximated by $2/(\sqrt{\tau_{po}/\tau_{no}} + \sqrt{\tau_{no}/\tau_{po}})$ for $V_{EB} > 4kT/q$ if the energy level of the recombination center is located nearly at the intrinsic Fermi level and if $\tau_{po} = \tau_{no}$.

The a. c. equations may be obtained from (1) and (2) in the usual way of expanding the currents and voltages in the form of $A = A_O + A_1 \exp(j\omega t)$ and assuming $A_O >> A_1$.

The Emitter-Base Surface Diode Characteristics

The theoretical analysis of the surface diode characteristics is extremely complicated since it involves the solution of three nonlinear simultaneous partial differential equations[5,6] in a two dimensional geometry. The problem is further complicated by the discontinuity of the crystal lattice at the Si-SiO_2 interface which may result in high concentration of surface states that are continuously distributed in the energy gap. Preferential precipation of impurities, such as gold or other metals at the interface might occur due to the high electric field near the surface. These impurities may provide additional recombination sites at the interface which have surface recombination properties different from those in the bulk. The derivation of the surface diode current-voltage relation is possible only by using rather simplified assumptions based on plausible physical considerations and experimental observations.

The observed current-voltage relation of the emitter-base surface diode, SD, shown in Fig. 9(b), may be approximated by

$$J_{SD} = j_{SD}\, 2\sinh(qV_{EB}/mkT) \quad (3)$$

where both j_{SD} and m are functions of the grid-base or grid-emitter voltage. The factor m may vary from 1 to as high as 10 or more depending on the width and doping of the surface channel, and it also varies with the emitter-base voltage. The usually observed values of m for the emitter-base surface diode of the tetrodes such as that in Fig. 11 are in the range of 1.5 to 4, and the coefficient j_{SD} has a thermal activation energy of from half of the energy gap to one-fourth of the energy gap in silicon.

The experimental data given in Fig. 12 and 13 where VGE is plotted against I_{BE} for two oxide thicknesses indicate that the current

flowing in the surface diodes may result from two two distinct mechanisms. Fig. 12 shows that 1NP2 has only one type of behavior which is characterized by the appearance of a maximum in the diode current versus the grid voltage characteristics. The data of the high grid voltage tetrode, 1W1-3, shown in Fig. 13 appears to have also only one type of behavior which is characterized by a monotonically increasing current with increasing or decreasing grid voltage from the minimum current point. The data for 2L2-4 given in Fig. 12 clearly shows both types of behavior indicating that both components of the surface current are present and are superimposed on each other. The fine structures near V_{GE}=80v in the characteristics of 1W1-3 in Fig. 13 suggest that a second component of surface current is probably present.

For mathematical simplicity, ease of discussion, and physical understanding, it is convenient to break up the surface into two regions: (1) the region corresponding to the projection of the emitter-base transition region onto the surface; and (2) the channel region outside of the emitter-base transition region. In the transition region, the quasi-Fermi levels are nearly horizontal since most of the current is from space charge generation and recombination. In the channel region there is considerable voltage drop. In the following discussion, it will be shown that the surface current component which has a maximum versus the grid voltage may be accounted for by the recombination current in the transition region on the surface; while the surface current component which is monotonically increasing from a minimum value with increasing or decreasing grid voltage corresponds to the recombination current flowing in the surface channels which are formed by the grid voltage outside of the emitter-base transition region.

Surface Recombination Current In The Transition Region

The diode current due to surface recombination in the transition region would be expected to go through a maximum when the surface potential (or the intrinsic Fermi level at the surface) is moved by the grid voltage to within (1/2) $\log_e(S_{po}/S_{no})$ from the average value of the quasi Fermi levels for electrons and holes where S_{po} and S_{no} are the surface recombination velocities for holes and electrons on degenerate n-type or p-type surface. The diode current from this region would decrease asymptotically to constant values when the grid voltage is either decreased or increased from the value of the maximum surface recombination. The shapesof the diode current versus grid voltage or surface potential characteristics are very similar to that of the surface recombination velocity observed on uniform surfaces which have no junctions. Several maxima may be expected if there are several independent surface recombination centers.

The derivation of this result may be developed using a simple example: the ideal symmetrical junction with linear potential variation in the junction. In Fig. 14, the energy band diagram at the surface for a symmetrical emitter-base surface diode is shown. The surface energy band diagram in Fig. 14(b) corresponds to the condition of zero grid voltage and forward bias of the junction. The diagram in Fig. 14(c) includes only the surface potential curves, u_S (dashed) and the quasi Fermi levels for electrons and holes on the surface. However, Fig. 14(c) shows the relative position of the surface potential, u_S under various grid voltages.

The steady-state recombination rate per unit surface area may be obtained from the Hall-Shockley-Read statistics.[7] It is given by

$$U_S = n_i\sqrt{S_{po}S_{no}}\,\sinh(qV_{BE}/2kT) \times \{\cosh[u_S-(u_p+u_n)/2 + (1/2)\log_e S_{no}/S_{po}] + \exp(-qV_{BE}/2kT)\cosh[u_s-u_t + (1/2)\log_e S_{no}/S_{po}]\}^{-1} \quad (4)$$

where S_{po} and S_{no} are the surface recombination velocities on degenerate n-type and p-type surfaces respectively, u_p and u_n are the hole and electron quasi Fermi levels, and u_s-u_t is the energy level of recombination centers measured from the intrinsic Fermi level. These energy levels are normalized to the thermal voltage, -kT/q. In Fig. 14 and Eq. (4), it is assumed that the change of quasi Fermi levels across the transition region is negligible but may be appreciable in the channel due to channel resistance. The maximum surface recombination occurs when the surface potential in (4) is such that

$$u_S = (u_p + u_n - \log_e S_{no}/S_{po})/2 \quad (5)$$

The position of the maximum surface recombination rate is shown in Fig. 14(c) as heavy dots. For this example, it is assumed that an accumulation layer is formed on the p-type surface under equilibrium or zero junction and grid voltage. The formation of an accumulation layer, for example, might arise due to the work function difference between the silicon, oxide and the metal of the grid electrode. Thus, the position of the maximum recombination rate, U_S(max.), coincides with the center of the junction when the grid voltage is a positive value labeled V_G = 1 in Fig. 14(c). At zero grid voltage, the point of maximum surface recombination is slightly to the right of the center of the junction or closer to the n-region. As the grid voltage is increased from large negative values, the maximum surface recombination point moves from the n-region into the junction transition region and then out of the transition region into the p-region.

The surface potential, u_S may be related to the grid voltage by an analytical procedure similar to that used for the metal-oxide-silicon capacitors.[8] However, the situation is somewhat different in the present case since there is a large minority carrier concentration near the surface which is injected by the forward biased junction. The results of the metal-oxide-semiconductor capacitor may be used as a zeroth approximation using the carrier concentration in the transition region instead of that in the bulk to obtain the surface potential from the grid voltage.

The surface current may be approximated by assuming a constant recombination rate on the surface of the transition layer. The result is

$$I_{SD} = q\, W_{ET}\, U_S\, L_S \qquad (6)$$

where L_S is the circumference, W_{ET} is the emitter base transition layer width in the bulk, and U_S is given by (4) evaluated at the center of the junction transition layer. Thus, using (4) and (5), (6) has the form given by (3) in which j_{SD} is a function of the grid-base voltage. Furthermore, (6) predicts the observed shape of the junction current versus grid voltage characteristics of device 1NP2 shown in Fig. 12. The maximum occurs at a grid voltage corresponding to the surface potential condition given by(5).

In the ideal symmetrical junction considered here, the shape of the I_{SD} versus V_G plot is symmetrical with respect to the grid voltage corresponding to point of maximum surface recombination or maximum current. In the real emitter-base surface diodes of the NPN surface-potential controlled transistors, the surface of the n-emitter is more heavily doped than the surface of the p-base. Thus, an unsymmetrical junction is usually obtained and the recombination rate would be asymmetrical with respect to the grid voltage of the maximum surface recombination current.

The diffusion current along the surface outside the transition region is similar to that in the bulk and is usually small compared with the recombination current in the transition region on the surface if a surface channel or an inversion layer is not formed.[6] The situation is more complicated if the width of the grid is comparable with the diffusion length.

In the case where the surface diffusion length is very long compared with the width of the grid electrode, the electron Fermi level u_n shown in Fig. 14(c) would be quite horizontal in the p-region under the grid, and the entire surface area under the grid electrode in the p-region would be important for carrier recombination, for positive grid voltage. An approximate result may be obtained by assuming no voltage drop along the surface under the grid electrode in the p-region. The expression for current is in the same form as that given by (6) if W_{ET}, the width of the emitter-base transition layer, is replaced by the half width of the grid electrode, $W_G/2$, and the recombination rate U_S is evaluated at the center of the region under the grid electrode on the p-side. A similar consideration applies to the n-region when the grid voltage is negative.

Channel Current

As the grid voltage is increased to certain value, such as that corresponding to $v_G = 3$, shown in Fig. 14(c), the surface potential is pulled down and coincides with the Fermi level of holes on the surface of the p-region outside the junction transition region under the metal grid electrode. The surface under the grid in the p-region now becomes intrinsic. Further increase of the grid voltage results in the formation of an inversion layer although not a channel on the p-surface, since the energy band is also lower in the n-region.

The appearance of a channel on one side of the junction under the grid where the band is flat or lower with respect to the other side of junction near the surface, such as that shown in Fig. 10, may occur in asymmetrical junctions. The requirement for channel formation is that the grid voltage can change the surface potential or bend the energy band perpendicular to the surface effectively only on one side of the junction and not on the other side. A channel will be formed on the side where the surface potential can be changed more readily by the grid voltage. The readiness of bending the band may be controlled by the thickness of the oxide, the impurity concentration at the surface and the surface states densities. In the case where a channel can be formed, it can occur at a grid voltage either before, at the same time or after the onset of surface inversion. For example, let us consider the case shown in Fig. 15 in which the surface potential on the n-side is locked in position and cannot be bent by the grid voltage appreciably. Then, as the grid voltage is increased, the surface potential or the energy band on the p-side will be pulled down and become flat with respect to the n-side ($V_G = 3$) before the surface potential, u_S, coincides with the Fermi level for holes, u_p ($V_G = 4$). Thus, a channel for electrons is formed on the p-surface before the onset of surface inversion, $u_S = u_p$.

In the practical silicon tetrodes fabricated, surface inversion and channel formation appears to be more readily achieved on the p-type surface than on the n-type surface under the oxide. This is evidenced by the experimental observations that the emitter-base surface diode current or the collector current of the tetrode, when plotted as a function of grid voltage, has almost the identical behavior for the NPN and the PNP tetrodes. Some typical data were shown in Fig. 5 and 6. Since the oxide thickness is fairly uniform and probably slightly thicker on the base-side than on the emitter side due to the diffusion sequence, the nonuniformity of oxide would probably not account for the same behavior of NPN

and PNP tetrode. The oxides on the emitter-base junction for the NPN and the PNP tetrode are grown in the same way so that it is unlikely that there is high concentration of surface states preferentially located on the n-region of the silicon-silicon oxide interface. The surface state density required to interpret the similarity of the transfer characteristics of NPN and PNP tetrode would be extremely high, in order that channel or surface inversion can barely be obtained in the n-type base surface of the PNP unit. However, it would still not explain why a channel is easily formed on the emitter surface of the PNP unit which is presumably doped to 10^{20} to 10^{21} surface concentration. Thus, it must be concluded that impurity segregation is important so that the n-type impurity is rejected by the oxide during thermal oxidation and forms a degenerate layer near the surface while the p-type impurity, in this case boron, prefers the silicon oxide under the thermal oxidation condition and thus is depleted at the silicon-silicon oxide interface. This conclusion is in agreement with the thermodynamic calculation of Thurmond.[9]

The formation of the channel or surface inversion layer in the base region or the emitter region may be calculated in a zeroth approximation by neglecting the electrons injected into the channel or inversion layer by the forward-biased emitter-base junction. This assumption is a somewhat better approximation here than it was applied previously to the surface within the transition region in calculating the surface recombination current. The results of the metal-oxide-semiconductor capacitor analysis can again be applied.[8]

The condition for the onset of surface inversion, i.e. u_S coincides the Fermi level for the majority carriers, is plotted in Fig. 16. Let us consider a numerical example using unit 1W1-3 shown in Fig. 13. The oxide thickness is approximately $L_o = 0.6$ microns and the surface concentrations of the impurity in the base and the emitter regions are approximately 10^{18}/cc and 10^{20}/cc. From Fig. 16, one may obtain an approximate effective width of the surface space charge layer width using the surface concentration instead of the true profile of the diffused emitter or base layers. In the base region, Fig. 16 gives $x_s = 0.02$ micron indicating the surface inversion layer is very thin and concentrated near the surface so that the use of the surface concentration instead of the true impurity concentration profile is a good approximation. From Fig. 16, the electric field in the oxide at the onset of surface inversion may also be obtained. In the base region, $E_o = 10^6$v/cm and the voltage drop in the oxide is $V_o = L_o E_o = 60$ volts. The voltage drop in the space charge region of the semiconductor is approximately 0.4 volts, given by $u_S - u_{SO}$ in Fig. 16. Thus, the total grid voltage required for surface inversion in the p-base is about 60 volts which corresponds well with the observed voltage where there is a rapid rise of I_{BE} shown in Fig. 13.

A similar calculation can be made to obtain the required grid voltage for surface inversion in the n-emitter. Fig. 16 does not cover the degenerate range, however, a linear extrapolation on the log-log plot may be used for a rough estimate which yields $E_o = 5 \times 10^7$v/cm or $V_o = 3000$ volts, a value considerably greater than the dielectric breakdown of the oxide.

In order to make a better approximation, it is necessary to include the injected carrier by the forward biased junction. In the first order approximation, one may replace the carrier concentration in the bulk, p_o(or n_o) by $n_i \exp(u_p - u_B) - n_i \exp(u_B - u_n)$ where u_p and u_n are the quasi Fermi potential for holes and electrons and u_B is the intrinsic Fermi potential in the bulk.

An exact analysis of the recombination current in the channel or the surface inversion layer is quite complicated since the channel or the inversion layer is quite thin and highly resistive. Consequently, there is a considerable amount of voltage drop along the length of the channel. For example, let us consider the case of an inversion layer and a channel formed at the same time in the p-region such as that shown in Fig. 10. As the surface inversion layer and channel are gradually formed by increasing the grid voltage, a transition region of high recombination rate, given by the condition similar to (5) if u_s is replaced by u , appears next to the surface under the grid and becomes rapidly effective for minority carrier recombination. The recombination rate decreases along the channel in the y-direction in Fig. 10 due to the voltage drop in the channel. Thus, in calculating the total current flowing into the channel due to carrier recombination in the channel, it is necessary to take into account the voltage drop along the channel.

The mathematical analysis of the channel current can be worked out with a procedure very similar to that used by McWhorter and Kingston[10] which was subsequently extended by Culter and Bath[11] to include forward bias on the junction for silicon alloyed diodes. These authors take into account the voltage drop in the channel. However, they neglected to use the appropriate recombination rate in the channel, which provides the most important voltage dependence of the channel current. For example, Culter and Bath took the diffusion current outside the channel to calculate the total current in the channel for silicon diodes. This approximation is valid only for channels in germanium diodes, since it is well-known that the most important current in silicon junctions below or near room temperature come from minority carrier recombination in the transition region of the junction.[6] Consequently, the results of Culter and Bath when correlated with experimental data in silicon[11] require a channel length which is considerably greater than the maximum physical dimension of the silicon wafer of the diode. Furthermore,

their theoretical results cannot account for a voltage dependence of $j_{SD}\exp(qV/mkT)$ of the channel current in which m is greater than 2 and j_{SD} has a thermal activation energy less than half of the energy gap. The extension of these analyses to the case appropriate to silicon junctions is fairly obvious if carrier recombination in the transition region in the channel is taken into consideration. The results can then account for a value of m as high as 4 and an activation energy of j_{SD} as low as 1/4 of the energy gap of silicon.

The calculation of the channel current may be formulated with reference to Figs. 10 and 15. It is assumed that the surface potential is changed by the grid voltage only in the p-region and not in the n-region so that a channel will definitely be formed with positive grid voltage and an equipotential surface similar to that given by Fig. 10 is obtained.

It is also assumed that the quasi Fermi level for holes, u_p, is flat and that there is negligible recombination near the surface of the channel so that the total current flow there is entirely carried by electron flowing into the channel from heavily doped n-type emitter. Thus, the hole and the electron currents in the channel are[4,5]

$$I_p(y) = -qD_p p(du_p/dy)\, W_I L_S = O \tag{7}$$

and

$$I_n(y) = -qD_n n(du_n/dy)\, W_I L_S \tag{8}$$

where D_n is the diffusion constant of electrons in the channel, n is the electron concentration in the channel, W_I is the inversion or channel width in the x-direction and may be a function of y, and L_S is the circumference. The total current in the channel at a distance y from the emitter-base junction may be obtained (using the continuity relation) and is given by

$$I = I_n(y) \tag{9}$$

It is further assumed that the electron concentration is a constant along the surface of the channel. Thus, the change of the electron quasi Fermi level, u_n, is entirely due to the voltage drop in the channel, i.e.

$$n_S = n_i \exp(u_S - u_n) = \text{constant} \tag{10A}$$

or

$$du_S = du_n \tag{10B}$$

where u_S is the surface potential normalized to the thermal voltage kT/q. The results given by (10A) and (10B) based on the assumption of constant electron concentration on the surface is consistent with the relation

$$u_p - u_n = v \tag{11}$$

where v is the normalized voltage drop along the channel. Thus, using (7), (8), (10B) and (11), and replacing n by n_S in (8), the total current flowing in the channel given by (9) may be written as

$$I(y) = qD_n n_S W_I L_S\,(dv/dy). \tag{12}$$

Another relationship between current and voltage along the channel may be derived by calculating the current flowing out of the channel in the x-direction across the surface potential barrier (see Fig. 10). This current, $J_x(V)$, is provided by the electron and hole recombination in the region near the surface where the condition of maximum recombination rate given by

$$u = (u_p + u_n - \log_e S_{no}/S_{po})/2 \tag{13}$$

holds. This condition is similar to that given by (5) for recombination in the transition region on the surface. The recombination current, $J_x(V)$, has the form given in reference 6. With the help of the continuity relation, div $\bar{I}$ = O, the change of the current flow along the channel in the y-direction may be set equal to the current flowing out of the channel in the x-direction due to electron hole recombination. Thus,

$$dI(y)/dy = L_S\, J_x(v) \tag{14}$$

where

$$J_x(v) = q\, W_I U_I \tag{15}$$

is the recombination current per unit area in the transition region of the surface barrier in the channel. The major variation of $J_x(v)$ with voltage v, for positive voltage or forward bias, comes from the term sinh(v/2) in the recombination rate expression for U_I which is similar in form to (4).[4,6] Thus, (15) may be written as

$$J_x(v) = j_{SI} 2\sinh v/2 \tag{16}$$

where

$$j_{SI} = qn_i W_I f(b)/2\alpha\sqrt{\tau_{po}\tau_{no}} \doteq qn_i W_I/(\tau_{po} + \tau_{no}). \tag{17}$$

Using (16) and (17), (14) may be written as

$$dI(y)/dy = [2qn_i W_I L_S/(\tau_{po} + \tau_{no})]\sinh v/2 \tag{18}$$

The two equations, (12) and (18) may be combined to yield

$$\frac{d^2v}{dy^2} = \frac{2n_i}{n_s D_n (\tau_{po} + \tau_{no})} \sinh\frac{v}{2} \tag{19}$$

if it is assumed that W_I is not a function of y. This assumption is valid inside the channel not too close to the edge of the emitter-base transition region. The solution of (19) may readily be obtained if the channel or the inversion layer is short compared with the width of the grid electrode. Under this condition, it may be assumed that at y = ∞, v(y) = 0, I(y) = 0 and the solution for the current flowing into the surface inversion layer or channel is

$$I_{SI} = I(0) = 4qW_I L_S \sqrt{\frac{D_n n_i n_s}{\tau_{po} + \tau_{no}}} \sinh\frac{qV_{EB}}{4kT} \tag{20}$$

Thus, the above result has again the desired form required by the empircal formula given by (3). In particular, (20) shows that m may be as high as 4 if surface channel or surface inversion layer current dominates. It also shows that the thermal activation energy of j_{SD} in (3) is that of $\sqrt{n_i}$ in (20) or one fourth the energy gap of silicon. It further shows that the dependence of the channel current on the grid voltage comes from j_{SD} or W_I and n_s. More accurate solution of (12) may be obtained from (19) if finite boundary conditions are used and the more accurate expression for $J_x(v)$ is used.

Small Signal Equivalent Circuit

The equivalent circuit of the tetrode was obtained in Fig. 9(b) from the physical circuit in Fig. 9(a). The most useful small signal equivalent circuit parameters are the hybrid parameters in the grounded emitter connection given by the following equations

$$V_1 = h_{11}I_1 + h_{12}V_2 + h_{13}V_3 \tag{21}$$

$$I_2 = h_{21}I_1 + h_{22}V_2 + h_{23}V_3 \tag{22}$$

$$I_3 = h_{31}I_1 + h_{32}V_2 + h_{33}V_3 \tag{23}$$

The numerials refer to the electrodes shown in Fig. 9(b): 1 = Base, 2 = Collector, 3 = Grid. Thus, $h_{21} = h_{fe}$, $h_{22} = h_{oe}$, $h_{23} = g_m$, $h_{12} = \mu_{12}$, etc. The hybrid circuit representation is preferred since it reduces to the usual transistor representation if the base lead is used as input and grid is either d.c. biased or modulated. If the base lead is r.f. open circuited and the grid is the input, then the hybrid equations reduce to those similar to the vacuum tube pentode and thus facilitates the comparison with vacuum pentodes.

The derivation of these h parameters from the equivalent circuit shown in Fig. 9(b) is straight forward with the help of (1) and (2) for the ideal transistor and (3) for the surface diode. The algebra involved is tedious and only two important parameters will be discussed in the open base operation. These are the transconductance, $h_{23} = g_m$, and the output admittance, $h_{22} = h_{oe}$. The complete expression, neglecting the base resistance r_b' and assuming open base, i.e., I_1 (Base) = 0, are:

$$g_m = h_{23} = h_{11}\left[j\omega C_{ge} + g_{SD}\sqrt{\frac{|I_c - I_{co}|}{j_D\alpha_T}}\right] \times \left[|I_c - I_{co}|\frac{q}{kT} - j\omega C_c\right] \tag{24}$$

$$Y_c = h_{22} = j\omega C_c + h_{11}(j\omega C_c + g_c) \times \left[|I_c - I_{co}|\frac{q}{kT} - j\omega C_c\right] \tag{25}$$

where

$$h_{11}^{-1} = \left[g_c + \sqrt{\frac{|I_c - I_{co}|}{j_D\alpha_T}}\, g_{SD} + h_{fe}^{-1}|I_c - I_{co}| \times \frac{q}{kT} + j\omega(C_c + C_e + C_{ge})\right] \tag{26}$$

$$h_{fe}^{-1} = \frac{1 - \alpha_T\gamma_N}{\alpha_T\gamma_N} + \frac{(j_{ET} + j_{SD})}{2\sqrt{j_D\alpha_T|I_c - I_{co}|}} \tag{27}$$

and

$$g_{SD} = \frac{\partial j_{SD}}{\partial V_{GB}},$$

which is the transconductance of the surface diode. In this derivation, it is assumed that m = 2 in the surface diode characteristics.

These relations may be simplified for low and intermediate frequencies. For practical tetrodes, most of the terms in h_{11}^{-1} given by (26), may be neglected except $h_{fe}^{-1}|I_c - I_{co}|\frac{q}{kT}$, thus,

$$h_{11}^{-1} \doteq h_{fe}^{-1}|I_{co} - I_c|\frac{q}{kT}. \tag{28}$$

The output admittance becomes

$$Y_c = h_{22} \doteq (1 + h_{fe})\, j\omega C_c + h_{fe}g_c \tag{29}$$

and the transconductance is

$$g_m = h_{23} \doteq \left[j\omega C_{ge} + g_{SD}\sqrt{\frac{|I_c - I_{co}|}{j_D\alpha_T}}\right] \times h_{fe}\left(1 - j\omega C_c \cdot \frac{kT}{q}|I_c - I_{co}|\right) \tag{30}$$

Usually $j_D\alpha_T = 10^{-13}$amp, $j_{SE} = 10^{-14}$amp, $C_c = 10$pf, $C_{ge} \sim 10$pf and $g_{SD}\left(\frac{\alpha_T\gamma_N}{1-\alpha_T\gamma_N}\right) \sim 10^{-10}$ mhos.
Thus, the 3 db frequency of g_m given in (30) is essentially that of h_{fe}, in agreement with that observed in Fig. 7.

The low frequency transconductance in the intermediate collector current range may be approximated by

$$g_m \doteq g_{SD}\sqrt{\frac{|I_c - I_{co}|}{j_D\alpha_T}}\,\frac{\alpha_T\gamma_N}{1 - \alpha_T\gamma_N} \tag{31}$$

Only the first term in h_{fe}^{-1} given by (27) is retained in (31). The square root dependence of the transconductance on the collector current is observed in tetrodes in the intermediate collector current range and a sample of the experimental data is shown in Fig. 7.

Concluding Remarks

A new semiconductor tetrode, the surface-potential controlled transistors, is described in this paper. This tetrode employs a field effect electrode as the grid to modulate the emitter-base surface diode characteristics of the transistor and thereby the h_{fe} characteristics. The

grid features an extremely high input impedance of greater than 10^{15} ohms, and a small capacitance of a few picofarads. Transconductance of 7000 micromhos has been achieved which is frequency limited by the h_{fe} cutoff.

The elementary theory of the tetrode transistor is worked out and the theory of the current-voltage characteristics of the emitter-base surface diode is developed in detail.

The development of the surface-potential transistor adds to the transistors and diodes a new family of semiconductor devices which has additional control terminals and provides additional flexibility and heretofore unobtainable electrical performance.

Acknowledgment

The author is deeply grateful for the efforts of a number of his collegues who made the surface-potential controlled transistors a reality. In particular, he is indebted to G. E. Moore and V. H. Grinich for numerous suggestions. He wishes to thank D. A. Tremere and P. S. Flint for fabricating the devices and H. Bogert for providing the high frequency data shown in Fig. 7.

References

1. J. A. Hoerni, "Planar transistors," presented at the Electron Devices Conference, October 27-28, 1960; and published as Fairchild Technical Paper TP-14.

2. W. Shockley, "A unipolar field effect transistor," Proc. IRE, vol. 40, pp. 1365-1376; November, 1952.

3. A. L. McWhorter, "1/f noise and germanium surface properties," in Semiconductor Physics, pp. 207-228; editor: R. H. Kingston, University of Pennsylvania Press, 1957; and "1/f noise and related surface effects in germanium," T. R. 295, Research Laboratory of Electronics, M.I.T.; May 20, 1955.

4. C. T. Sah, "Effect of electrons and holes on the transition layer characteristics of linearly graded p-n junctions," Proc. IRE, vol. 49, pp. 603-618; March, 1961.

5. W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors," B.S.T.J. vol. 28, pp. 436-489; July, 1949.

6. C. T. Sah, R. N. Noyce and W. Shockley, "Carrier generation and recombination in p-n junctions and p-n junction characteristics," Proc. IRE, vol. 45, pp. 1228-1243; September, 1957.

7. R. N. Hall, "Germanium rectifier characteristics," Phys. Rev. vol. 83, p. 228, July, 1951; "Electron-hole recombination in germanium," Phys. Rev. vol. 87, p. 387, July, 1952; W. Shockley and W. T. Read, Jr., "Statistics of recombinations of holes and electrons," Phys. Rev., vol. 87, pp. 835-842; September, 1952; and D. T. Stevenson and R. J. Keyes, "Measurements of the recombination velocity at germanium surfaces," Physica, vol. 20, pp. 1041-1046; November, 1954.

8. The problem of the potential variation in a metal-oxide-semiconductor sandwich was analyzed by its inventor, J. L. Moll, in "Variable capacitance with large capacity change," IRE Wescon Record, part 3, pp. 32-36, August 18, 1959, and has since been analyzed by a large number of authors recently. Our notations follow closely that of Kingston and Neustadter, "Calculations of the space charge, electric field and free carrier concentration at the surface of a semiconductor," Jl. Applied Physics, vol. 26, pp. 718-720; June, 1955; More recent papers on the analysis of the oxide capacitor are: R. Linder, "Semiconductor surface varactor," Second Interim Report on Microwave Solid State Devices, pp. 13-40, Bell Telephone Laboratories, (Contract DA 36-039 sc-85325), November 30, 1960; and D. R. Frankl, "Some effect of material parameters on the design of surface space-charge varactors," Solid State Electronics, vol. 2, pp. 71-76; January, 1961.

9. C. D. Thurmond, "Distribution coefficients of impurities distributed between Ge or Si crystals and ternary alloys or surface oxides," Proc. Metallurgical Society Conferences, vol. 5, Properties of Elemental and Compound Semiconductors, editor: H. C. Gatos, pp. 121-140, Interscience Publishers, 1960; and M. M. Atalla and E. Tannenbaum, "Impurity redistribution and junction formation in silicon by thermal oxidation," B.S.T.J. vol. 39, pp. 933-946; July, 1960.

10. A. L. McWhorter and R. H. Kingston, "Channels and excess reverse current in grown germanium p-n junction diodes," Proc. IRE, vol. 42, pp. 1376-1380; September, 1960.

11. M. Cutler and H. M. Bath, "Surface leakage current in silicon fused junction diodes," Proc. IRE, vol. 45, pp. 39-43; January, 1957.

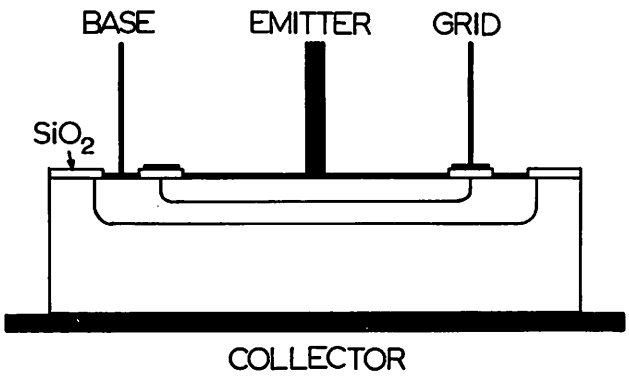

Fig. 1 A cross-sectional view of the surface-potential controlled tetrode.

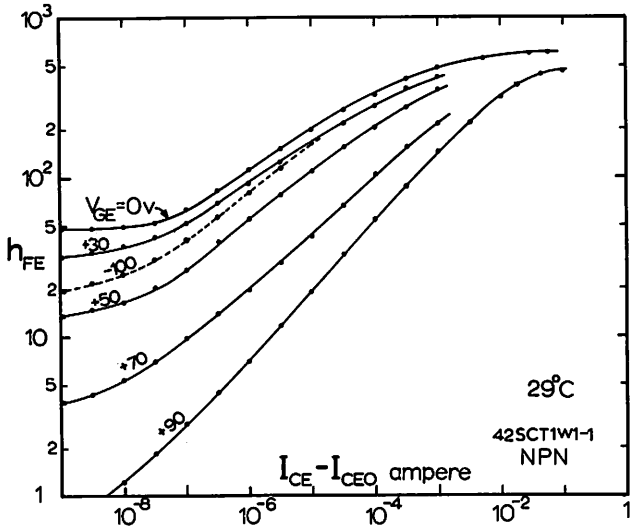

Fig. 2 The d.c. current gain, h_{FE}, for a NPN tetrode plotted as a function of collector current for several grid voltages.

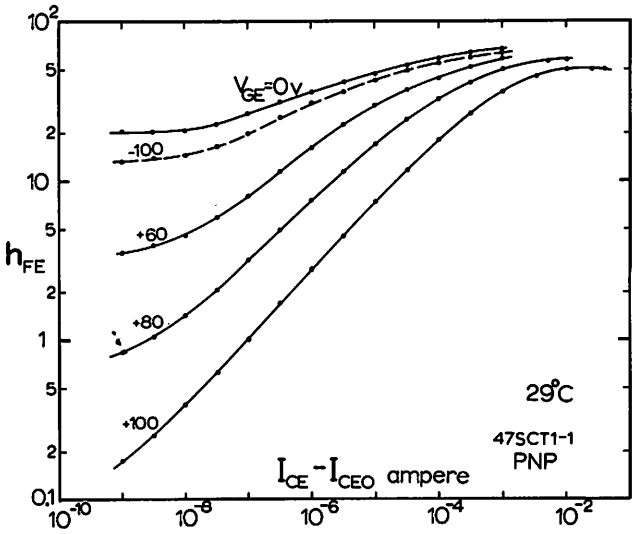

Fig. 3 The d.c. current gain, h_{FE}, for a PNP tetrode plotted as a function of collector current for several grid voltages.

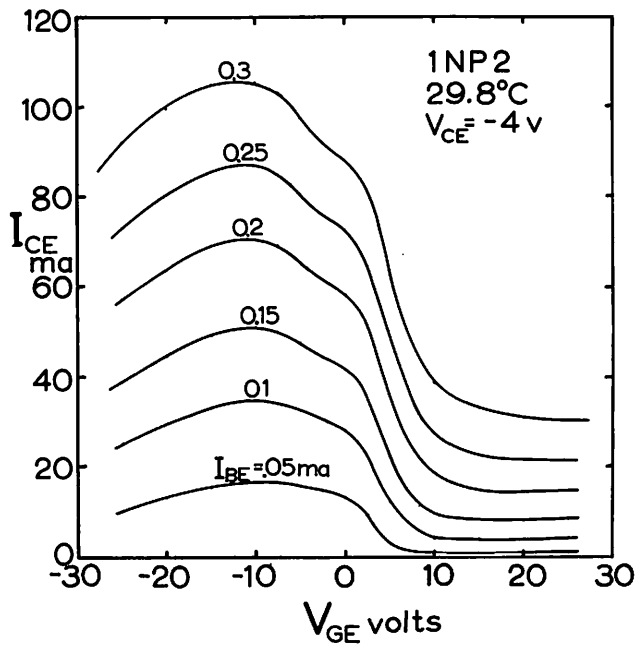

Fig. 4 The high current transfer characteristics of a NPN tetrode with thin oxide on the emitter-base junction.

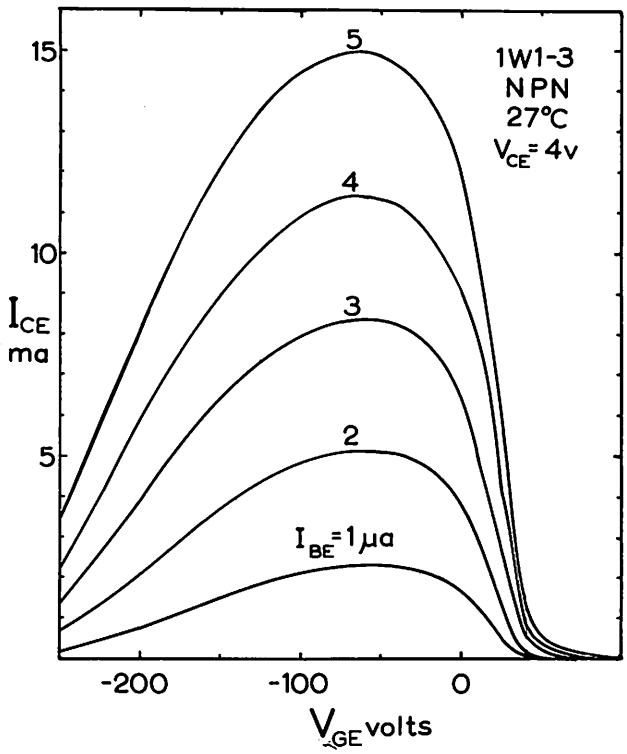

Fig. 5 The low current transfer characteristics of a NPN tetrode with thick oxide on the emitter-base junction.

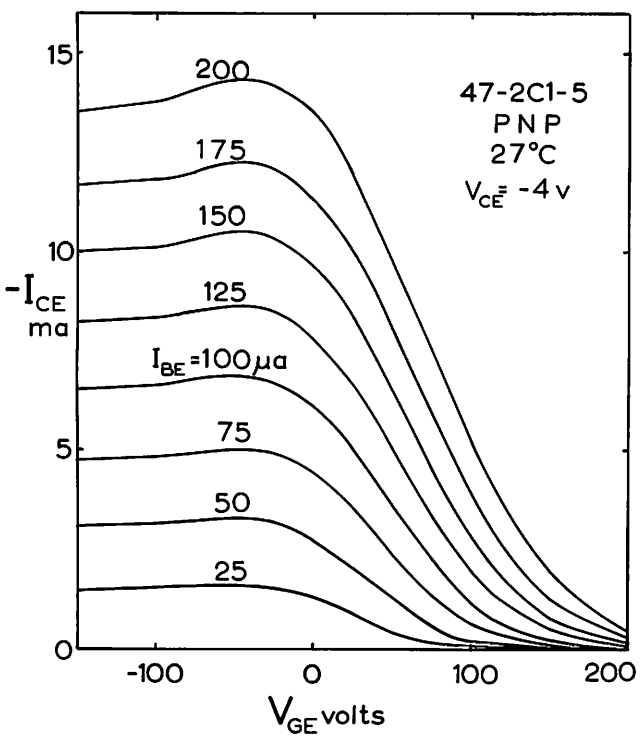

Fig. 6 The high current transfer characteristics of a PNP tetrode with thick oxide on the emitter-base junction.

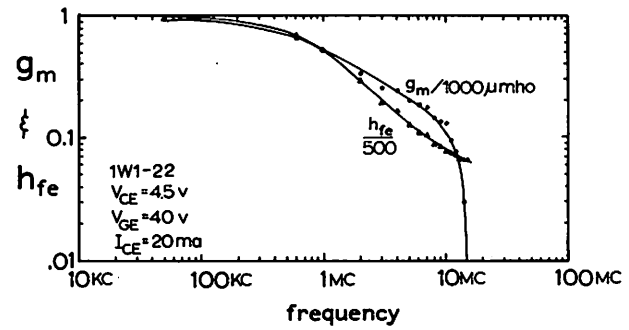

Fig. 7 The frequency dependence of the transconductance of a thick oxide NPN tetrode.

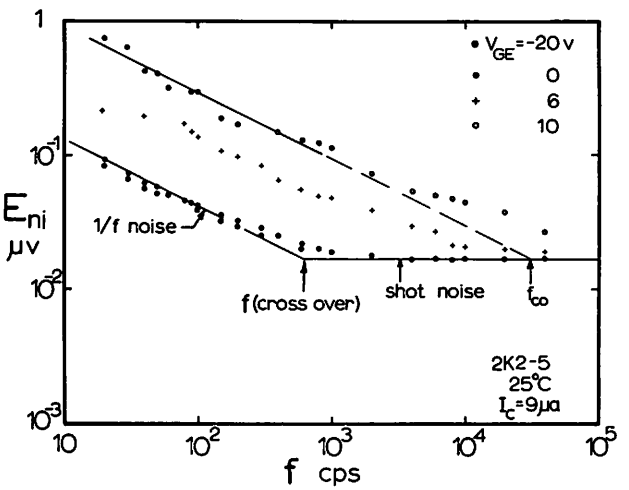

Fig. 8 The frequency dependence of base input noise voltage for several grid voltages.

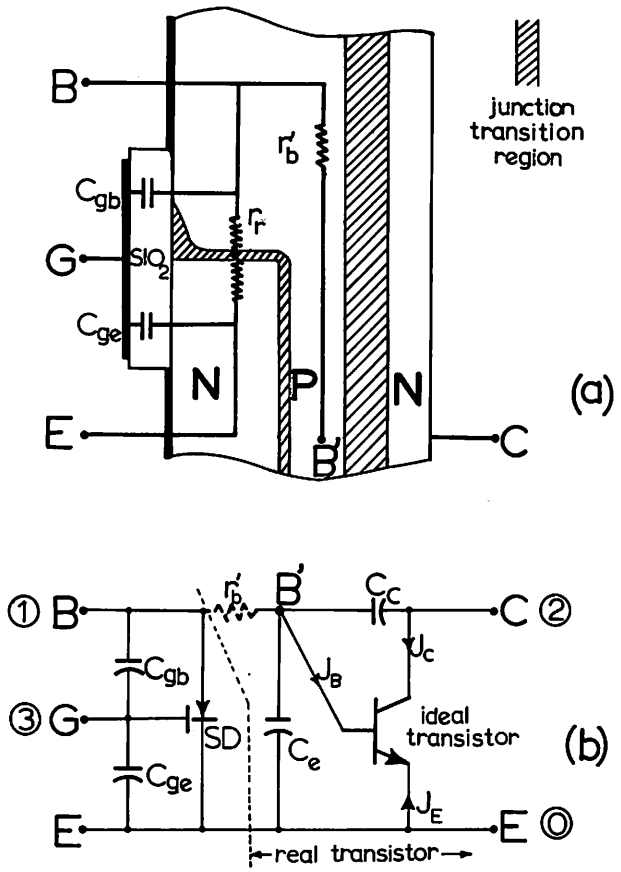

Fig. 9 (a) The physical circuit, and

(b) The equivalent circuit of the surface-potential controlled transistor tetrode.

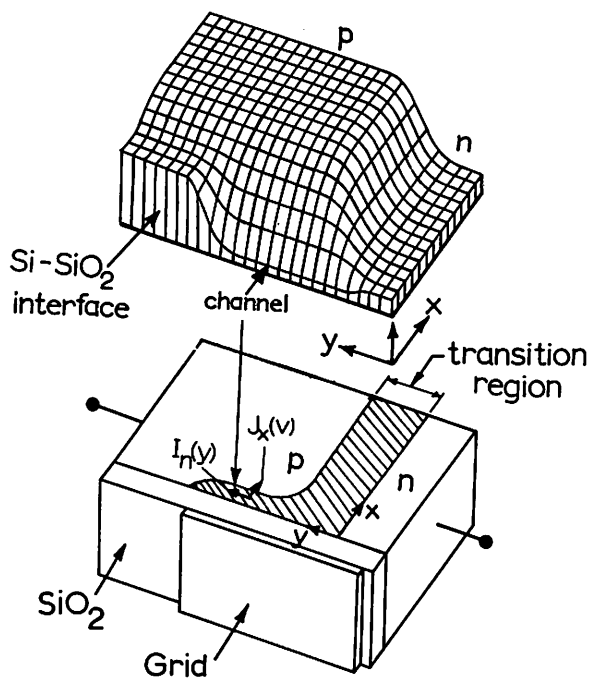

Fig. 10 The equipotential surface (top figure); and the cross-sectional view of an asymmetrical emitter-base surface diode in which the n-type emitter is more heavily doped than the p-type base.

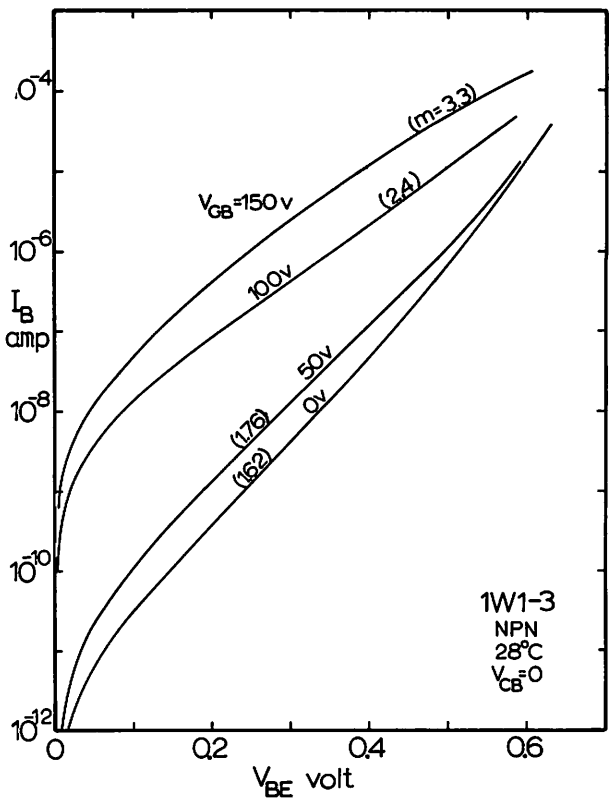

Fig. 11 The forward characteristics of an emitter-base surface diode with the grid voltage as a parameter. The quantity m is given by eq. (3).

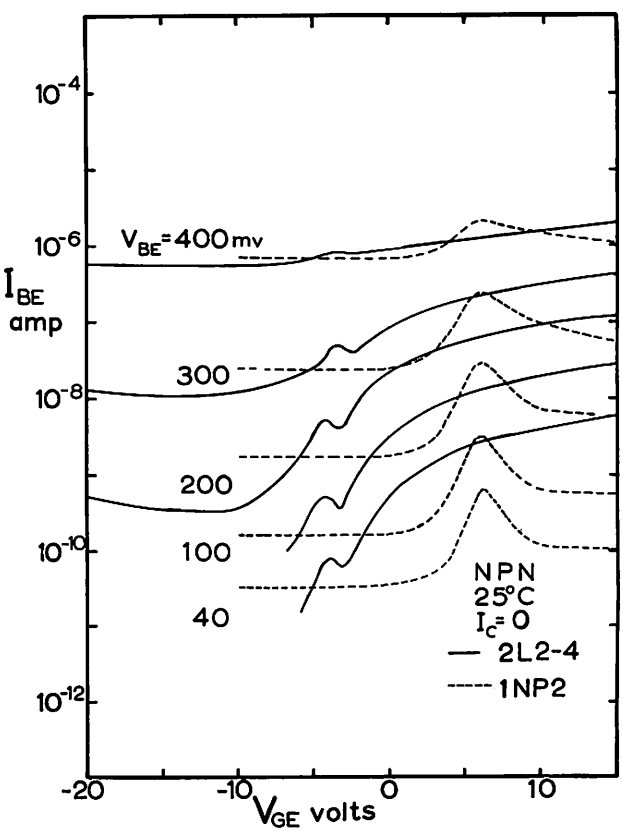

Fig. 12 The transfer characteristics of thin oxide emitter-base surface diodes with emitter-base voltage as a parameter.

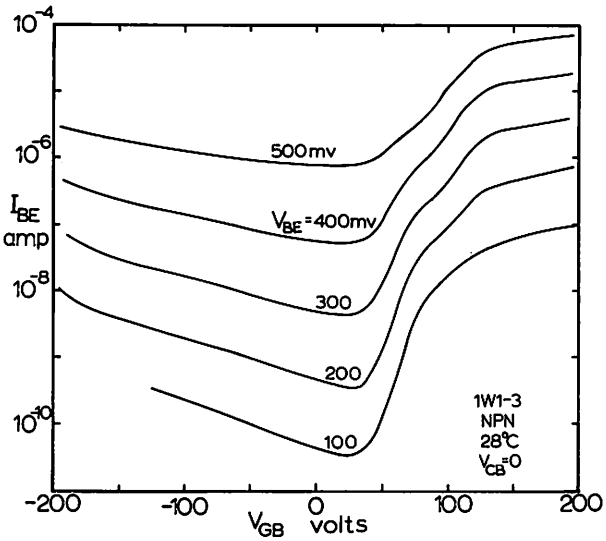

Fig. 13 The transfer characteristics of a thick oxide emitter-base surface diode with emitter-base voltage as a parameter.

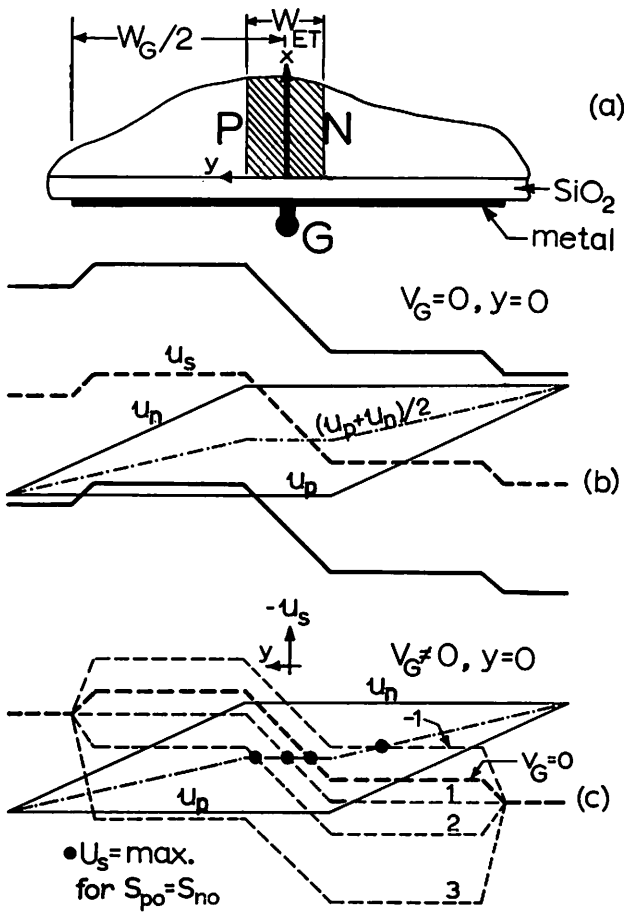

Fig. 14 (a) The cross-sectional view,
(b) The energy band diagram at equilibrium, and
(c) The surface potential variation along the surface of a hypothetical symmetrical emitter-base surface diode.

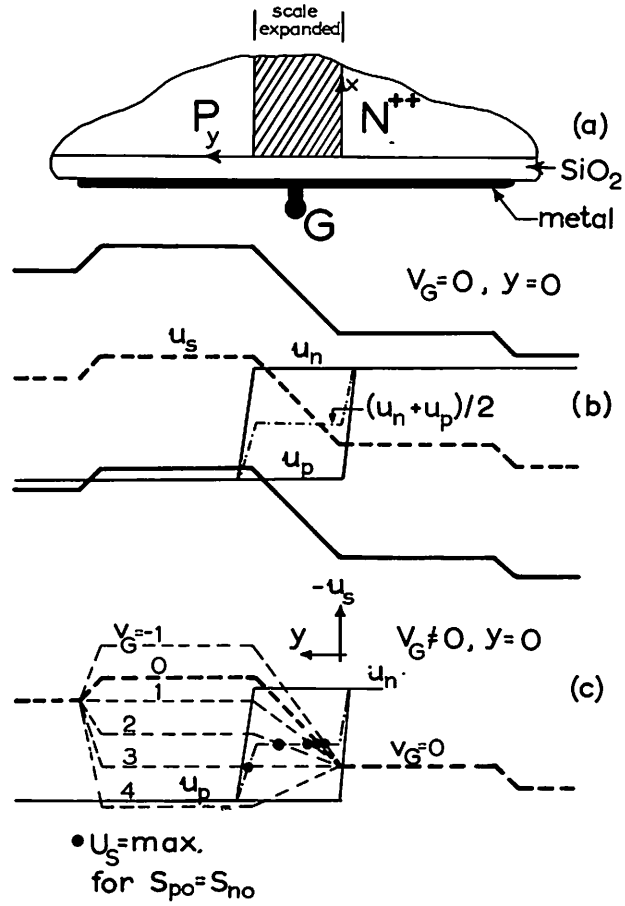

Fig. 15 (a) The cross-sectional view,
(b) The energy band diagram at equilibrium, and
(c) The surface potential variation along the surface of a hypothetical asymmetrical emitter-base surface diode.

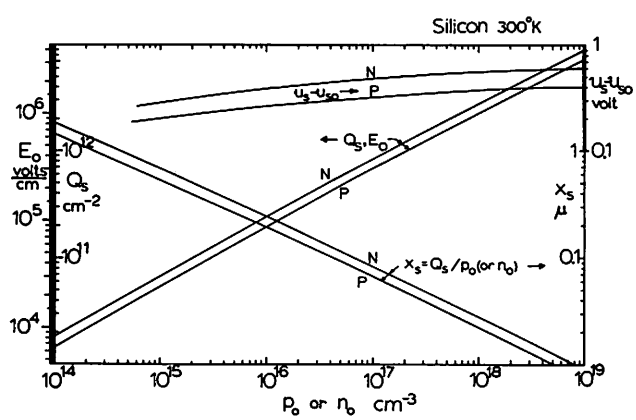

Fig. 16 A chart of the electrical field in the oxide, E_o; the total charge per unit area, Q_s; the equivalent thickness of the space charge layer in the silicon; and the voltage drop in the silicon plotted as a function of the carrier concentration in the bulk for both n and p-type silicon under the condition of surface inversion.

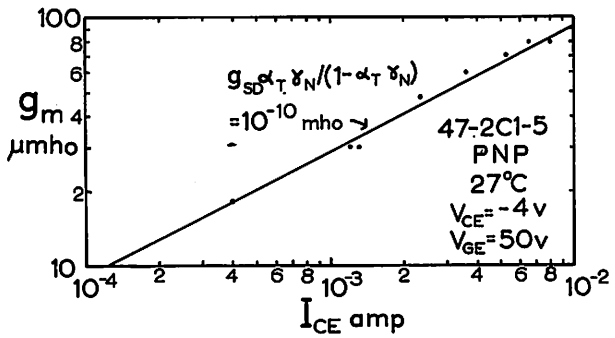

Fig. 17 The low frequency transconductance plotted as a function of collector current for a PNP tetrode.